

**METHOD, APPARATUS, AND COMPUTER
PROGRAM PRODUCT FOR AUTOMATIC
REMIX AND SUMMARY CREATION USING
CROWD-SOURCED INTELLIGENCE**

TECHNOLOGICAL FIELD

[0001] Example embodiments of the present invention relate generally to automated media generation and, more particularly, to a method, apparatus, and computer program product for utilizing crowd-sourced intelligence to automatically create remixes and summaries of events.

BACKGROUND

[0002] The use of image capturing devices has become prevalent in recent years as a variety of mobile devices, such as cellular telephones, video recorders, and other devices having cameras or other image capturing devices have become standard personal accessories. As such, it has become common for a plurality of people who are attending an event to separately capture video of the event. For example, multiple people at a sporting event, a concert, a theater performance or the like may capture video of the performers. Although each of these people may capture video of the same event, the video captured by each person may be somewhat different. For instance, the video captured by each person may be from a different angle or perspective and/or from a different distance relative to the playing field, the stage, or the like. Additionally or alternatively, the video captured by each person may focus upon different performers or different combinations of the performers.

[0003] Accordingly, it may be desirable to mix the videos captured by different people. However, efforts to mix the videos captured by a number of different people of the same event have proven to be challenging, particularly in instances in which the people who are capturing the video are unconstrained in regards to their relative position to the performers and in regards to the performers who are in the field of view of the videos.

[0004] The content capturing capabilities of mobile devices have improved much more quickly than network bandwidth, connection speed, and geographical distribution. Accordingly, there is great value to an end user if video can be recorded and value added content generated without the need for uploading, from a mobile device, large amounts of data, which is inherent to video recording. Some work has been done to generate panoramic views of events using ultra-high resolution video capturing equipment arranged contiguously to create a 360 degree view coverage of a venue (e.g., the FASCINATE project). This work has become possible due to the leaps in the media capture and network capabilities.

[0005] However, capitalizing on the ability to capture ultra-high resolution video using a thin client mobile device requires overcoming several hurdles. The biggest problem is that because bandwidth has not increased at a similar rate as video capturing capabilities, uploading high quality video content for generating value added content like remixes, summaries, etc. can often be impractical. In addition, however, the disparity in the media capture quality of recording devices and the potential absence of users in key spots on the field may result in gaps in event coverage (both spatial and temporal).

[0006] Finally, even in conjunction with an ultra-high resolution contiguous video capturing system, another problem is an inability to automatically find out an appropriate view

selection for a remix or summary of an event. In this regard, the main problem is related to determining the most relevant and interesting parts that should be included in a particular representation (based on the selection of a view) of the event, since most of the commonly available viewing apparatus will not match the dimensions, resolution, or connectivity to view the complete recorded content (i.e. the 360 degree view). First, for viewing the high resolution panoramic video content, a very high resolution display of large size is needed which is not readily available. Second, the network bandwidth needed to support the transmission of such high bit rate is also not readily available. Prior art systems have a drawback in that the intelligence for view selection is limited to single user's choice. Accordingly, there is a need to generate a more representative remix and/or summary of an event that takes into account the viewing preferences of an entire crowd.

BRIEF SUMMARY

[0007] Accordingly, a method, apparatus, and computer program product are provided to utilize crowd-sourced intelligence to automatically create remixes and summaries of events. In this regard, a method, apparatus and computer program product are provided to collect sensor and context data from a variety of thin client devices for use in automatic remix creation.

[0008] In a first example embodiment, a method is provided that includes receiving sensor and context data from at least one device, causing, by a processor, generation of a media remix based on the sensor and context data received from the at least one device, and causing transmission of the media remix to a client device. In this regard, the sensor data from the at least one device comprises at least one selected from the group consisting of: orientation with respect to north; orientation with respect to horizontal; position in three dimensional space; global positioning system (GPS) data; or location data, and the context data from the at least one device enables calculation of the depth of focus of the at least one device. Moreover, causing generation of the media remix may further be based on the sensor and context data of the client device.

[0009] In some embodiments, generation of the media remix includes identifying at least one focus of interest based on the sensor and context data, extracting relevant media segments from a recording engine based on candidate views corresponding to the at least one focus of interest, and generating the media remix based on the relevant media segments. In one such embodiment, identifying the at least one focus of interest based on the sensor and context data includes determining a location, orientation, and area of focus of the at least one device based on the sensor and context data, and identifying the at least one focus of interest based on the location, orientation, and area of focus of the at least one device. In another such embodiment, generation of the media remix further includes identifying the candidate views corresponding to the at least one focus of interest by evaluating candidate views from the recording engine based on at least one of: a comparison of distance of focus of the candidate view to distance of focus of the focus of interest, a comparison of an orientation of the candidate view with respect to the focus of interest, and detectability of the focus of interest in the candidate view using object detection or object recognition analysis; and selecting candidate views from the record-